Tetrahymena H4 genes: structure, evolution and organization in macro- and micronuclei

Gary A.Bannon, Josephine K.Bowen, Meng-Chao Yao* and Martin A.Gorovsky

Department of Biology, University of Rochester, Rochester, NY 14627, and *Department of Biology, Washington University, St. Louis, MO 63130, USA

Received 19 October 1983; Accepted 20 January 1984

ABSTRACT

The ciliated protozoan Tetrahymena thermophila contains two types of H4 histone genes (H4-I and H4-II). Southern blotting and analysis of DNA from nullisomic strains indicate that H4-I and H4-II are on different chromosomes and that only H4-II is closely linked to an H3 gene. No DNA sequence rearrangements are observed for either of the H4 genes when the transcriptionally inert, germ line, micronucleus is compared to the transcriptionally active, somatic macronucleus. Comparison of the H4-I gene and its flanking sequences to H4 gene sequences of other organisms indicates that there are evolutionary constraints on coding nucleotides that are unrelated to their protein coding function and that these evolutionary pressures operate at the level of translation.

INTRODUCTION

Although genes coding for histone proteins have been described in detail for a number of eukaryotes, these descriptions have been restricted largely to the animal kingdom. Within this kingdom, no single histone gene arrangement appears to have been conserved (1). Often, however, histone genes are linked. The first descriptions of histone genes (those expressed in early sea urchin development and in Drosophila) indicated that genes for all five histones were linked to form a unit which was tandemly repeated hundreds of times (2-10). More recent studies have indicated that histone genes in multicellular animals can be organized into clusters containing more than five histone genes (Xenopus; 11-15, chicken; 16,17) or may not be clustered at all (mouse; 18). In one case (Notophthalmus; 19), histone genes are clustered and repeated, but the clusters are separated by up to 50 kb of satellite DNA. Even in a single animal species different sets of histone genes can have different organizations. Thus, sea urchin early genes are clustered and highly repeated in a tandem array while genes expressed late in development are less redundant and are highly dispersed (20).

The only member of the fungal kingdom whose histone gene organization has

been described in detail is the yeast, Saccharomyces cerevisiae, where genes coding for the core histones (no H1 gene or H1 protein has yet been found) are arranged as four unlinked sets of gene pairs. Each set is composed of an H2A-H2B pair or an H3-H4 pair (21,22).

Plant histone gene organization has only been described in wheat. Tabata et al. (23) have shown the H4 gene to be reiterated 100 to 125 times per hexaploid genome and have described one recombinant clone in which the H4 and H3 genes are linked.

Information on histone genes in the protistan kingdom is limited to one study of macronuclear DNA in the ciliated protozoan, Stylonichia mytilus (24) which indicated that histone genes in this organism were completely unclustered. However, since the macronuclear genome in Stylonichia, as in other hypotrich ciliates, is fragmented into numerous DNA pieces each of which probably contains a single gene, this organization is not unexpected. No data are available on the arrangement of histone genes in Stylonichia micronuclei.

Because of the paucity of data on histone gene organization in protists and because previous studies indicated that Tetrahymena has the most unusual H4 amino acid sequence (25,26), we undertook to characterize the organization of histone H4 genes in Tetrahymena. We were particularly interested in comparing histone gene organization in Tetrahymena to that recently described in yeast since both organisms are rapidly dividing unicellular eukaryotes with long, independent evolutionary histories, presumably distinct from those leading to higher organisms. Since recent studies have emphasized the role of genome reorganization in the development of the somatic, transcriptionally active macronucleus from the germinal, transcriptionally inert micronucleus, we have also compared the organization of histone H4 genes in these two nuclei.

MATERIALS AND METHODS

Cells and Culture Conditions

Tetrahymena thermophila (strain B-1868-VII) were grown axenically in enriched proteose peptone (27) at 28°C.

DNA Isolation

Macro- and micronuclei from T. thermophila were isolated as described by Gorovsky et al. (27). DNA was isolated according to the methods of Bannon et al. (28).

Yeast Histone Probes

The yeast probes used in this study were isolated from pMS191 (22)

containing one copy of the yeast H4 and H3 genes cloned into the HindIII site of pSC101. Purification of the H4 and H3 coding regions used as probes has been described elsewhere (28).

Hybridizations

DNA probes were nick translated (29) to a specific activity of ~10^8cpm/µg DNA and hybridized to filters at 50°C (heterologous probes) or 65°C (homologous probes). After incubation for 12 hours, the filters were washed, dried, and put up for autoradiography against pre-flashed Kodak XAR-5 x-ray film with one intensifying screen at -80°C.

Southern Blots and DNA Restriction Digests

Transfer of DNA from agarose gels to nitrocellulose was accomplished according to published procedures (30). DNA restriction digests were done at 37°C in 50 mM NaCl, 6 mM Tris pH 7.4, 6 mM $MgCl_2$, 6 mM β mercaptoethanol, and 1 mg/ml BSA.

Cloning of Tetrahymena Histone Genes

A micronuclear library constructed by Yao (31) was screened by the plaque hybridization method (32) for clones containing H4 sequences. Approximately 4 x 10^4 pfu were screened for sequences containing H4 genes. Ten H4 containing clones (λGB 501-510) were obtained and further purified by two more rounds of hybridization. Propagation of phage was done on agar plates or in liquid culture in the host E. coli strain Q358 (r_k^- m_k^+ Su_{II}^+ 80^R) as described by Blattner et al. (33). Insert DNA from λGB508 was transferred from charon 4A to plasmid RVIIΔ7 (34) by directional cloning procedures described by Maniatis et al. (35).

DNA Sequencing Strategy

Sequencing was done according to the methods of Sanger et al. (36) using a series of clones constructed according to the methods of Hong (37). Briefly, a 2.1 Kb Sau 3A-Sau 3A Tetrahymena DNA fragment, which was positive for yeast H4 hybridization, was cloned into the BAM H1 site of the sequencing vector M13 mp8 (38). This subclone was then shortened by DNAse I digestion followed by cleavage at a HindIII restriction site lying between the universal primer and the Bam H1 site of M13mp8. DNA ends were then filled in with Klenow fragment and blunt end ligated.

RESULTS

Organization of H4 genes in macro- and micronuclei

We have recently presented evidence indicating that probes derived from yeast H3 and H4 genes recognize distinct mRNA sequences whose abundances

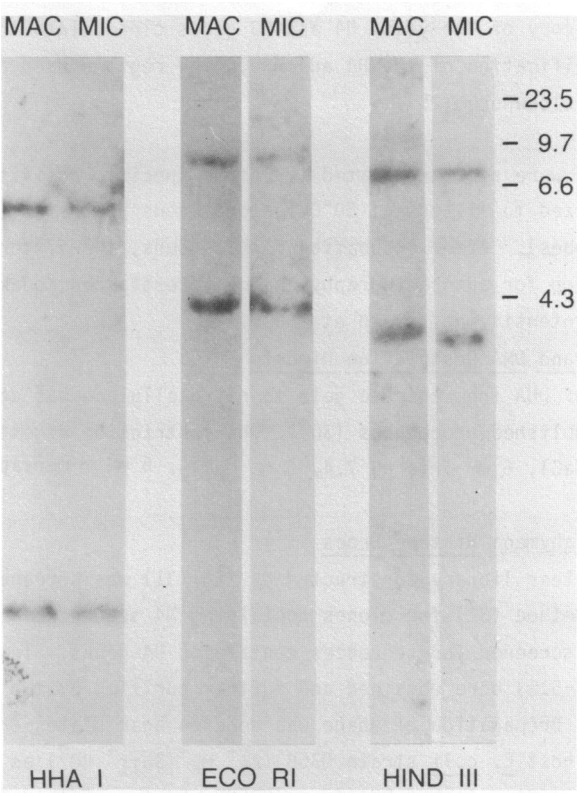

Figure 1. Sequence flanking Tetrahymena histone H4 genes are indistinguishable in macro- and micronuclei.
Macro- or micronuclear DNA was digested with Eco·RI, HindIII or Hha I, fractionated on 1% agarose gels, transferred to nitrocellulose filters and hybridized with the yeast H4 probe.

differ in growing and starved Tetrahymena (28). These studies indicated that these yeast sequences could be used as reliable probes for Tetrahymena histone genes. When Tetrahymena macronuclear DNA is digested with restriction endonuclease Eco·RI, HindIII, or HhaI and analyzed on Southern blots, two fragments show homology to the yeast H4 probe, suggesting that there are two H4 genes (designated H4-I and H4-II) in the Tetrahymena macronuclear genome (28). Since gene rearrangement is a prominent feature of macronuclear development in Tetrahymena (39,40), we wished to determine whether Tetrahymena H4 genes had identical germ-line (micronuclear) and somatic (macronuclear) configurations. Macro- and micronuclear DNAs were digested with Eco·RI, HhaI or HindIII and analyzed by Southern blotting using yeast H4 sequences as

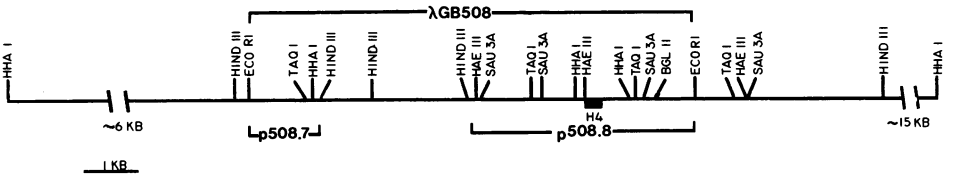

Figure 2. Map of sequences flanking the H4-I gene in macronuclear DNA. The original 8.6 kb insert (λGB508) containing the Tetrahymena H4 coding region was digested with Eco·RI and Hind III and subcloned into a plasmid vector digested with the same enzymes. The subclone containing the H4 gene (p508.8) was then mapped further and the coding region localized to a ~1 kb Hha I fragment using the yeast H4 probe. Flanking sequences ~20 kb in one direction and ~12 kb in the other direction were mapped by hybridizing end fragments from λGB508 (RI-Bgl II and p508.7) to macronuclear DNA digests.

probes (Figure 1). No rearrangements were detected near the H4 genes. From mapping data described below, it is clear that there are no detectable alterations of the DNA sequences ~5kb to the left and ~6kb to the right of the H4-I gene. Similarly, there are no detectable alterations in sequences flanking the H4-II gene, but it is not yet possible to determine the extent of the sequences flanking this gene which have been examined.

Isolation of a Tetrahymena DNA Fragment Homologous to the Yeast H4 Genes

Since H4 genes (and their flanking sequences) appear to be the same in

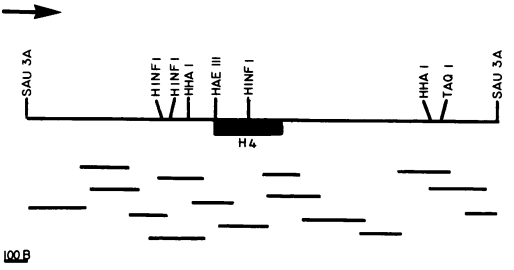

Figure 3. DNA Sequencing Strategy. The Sau 3A fragment of p508.8 containing the H4 coding region was subcloned into the Bam HI site of the sequencing vector M13mp8 (38). A mini-library of clones with different size inserts was then prepared according to the methods of Hong (37). The lines underneath the restriction map of the full length clone represent the portions of individually isolated clones which were sequenced. The length of the line indicates the number of bases we were able to determine from each clone. The box on the full length clone delineates the H4 coding region. The arrow indicates the direction of sequencing and direction of H4 gene transcription. For clarity, only 15 overlapping clones are shown in this figure. Most of the 15 regions shown were sequenced in two independent clones.

```
   1  ATTTATTTAT TTTTTTATTT CATTATTTAT TGTGTATTAT GATTTAGATA TATTTAATAA
  61  ATTATTATCA AATCTTTGTT GCTATGAAAA TTATTGGTTT TAATTAAATA TTTAAAAATT
 121  ATTTCAATAT AAACTGAACA TTTTAAAAAT AAAAATATCT CAAAAAAATT ATTTTTTATA
 181  AACATAACAT ATTTTAAATT TTTGAAAGTT ACTTTATGAA ATTAAAATAT AATTTAAAAA
 241  TAAAATTAAA ATAATACAAA CATGACTTTT TAATAATAAT AATAATTTTA TTTATTAATT
 301  TTAGAATATT TATTTTTTTT ATTTGATTTA ACAAATATAT TTTATAATAA AAAAAAATTT
 361  TAAATAAAAA TTTATTAATT TATTTTGAAT GAACTTAATG AATGTGTTTT TATTATTAAA
 421  TTTGCTATCA AATATATTTT AAATGCTCAA ACTAGATATT AGTCTATTAT CAATTGTAGT
 481  TATAGATAGT TAAATAGATT TTAGGGTTTT AAATTTTGAG GCTAAAATCC AAAAATGGTT
 541  GCATCTGGTG AGATATCTTC AAAGTATGGA TTAATTATTT CAAATTATTA GAAGGTAATT
 601  AATCTGCATA AATTCAAAAC TATAAAAATA AAACATTAAA ATTAATTCAA CCTTATTGAA
 661  GCATCAAAAT CTGAATCTCT AGAAAGACTG ATTCTGATTG GATAATTTTT CGGCGCTAAG
 721  GATTTTGGAT TAAAGAAAAT TAGATTTAAT TATTAATCAT GATTTGAATA GGATAGCAAG
 781  AATATTTGTT TGGTTTAAAA GGGAAAGCGG GTAATTATCA AAAATTTATA AATAATTTTA
 841  AAACAATAAA TAGAAAAACA AATAAGATTA TAAAAACTTA CAAAA

  890           900           910           920           930           940
   *             *             *             *             *             *
ATG GCC GGT GGT AAA GGT GGT AAA GGT ATG GGT AAA GTC GGA GCC AAG AGA CAC TCC AGA
Met Ala Gly Gly Lys Gly Gly Lys Gly Met Gly Lys Val Gly Ala Lys Arg His Ser Arg

  950           960           970           980           990          1000
   *             *             *             *             *             *
AAG TCT AAC AAG GCT TCC ATT GAA GGT ATT ACT AAG CCC GCT ATC AGA AGA TTA GCT AGA
Lys Ser Asn Lys Ala Ser Ile Glu Gly Ile Thr Lys Pro Ala Ile Arg Arg Leu Ala Arg

 1010          1020          1030          1040          1050          1060
   *             *             *             *             *             *
AGA GGT GGT GTT AAG AGA ATT TCC TCT TTC ATT TAC GAC GAC TCC AGA CAA GTC TTG AAG
Arg Gly Gly Val Lys Arg Ile Ser Ser Phe Ile Tyr Asp Asp Ser Arg Gln Val Leu Lys

 1070          1080          1090          1100          1110          1120
   *             *             *             *             *             *
TCT TTC TTA GAA AAC GTT GTT AGA GAC GCT GTC ACT TAC ACT GAA CAC GCT AGA AGA AAA
Ser Phe Leu Glu Asn Val Val Arg Asp Ala Val Thr Tyr Thr Glu His Ala Arg Arg Lys

 1130          1140          1150          1160          1170          1180
   *             *             *             *             *             *
ACC GTC ACT GCT ATG GAC GTT GTC TAC GCC CTC AAG AGA CAA GGC AGA ACT CTC TAT GGT
Thr Val Thr Ala Met Asp Val Val Tyr Ala Leu Lys Arg Gln Gly Arg Thr Leu Tyr Gly

 1190
   *
TTC GGT GGT TGA
Phe Gly Gly ---

1198  ACAAAATATT TATCTTAAAA AATTAAAAAG TAAAAAGCTG CATGCTTACT CAAAGGTAAT
1258  AGTGTAATTA TCTAGTTCTT TTATCTGAGA GAGTATGCTT TTTTCTATCG AGTGTTAGTG
1318  TAGCAATTTT CTAAAGTGCA TTGAGAGATT GAGnCAGAAA TGTTTAGAaC TTATTCACAt
1378  cAAATTAACT TAAGAAAAAT AATAACTTAC TTAATCAATT CACAATTAGC CATTATGAAT
1438  AACTAAACTA AAACATAAAG CTAATTTATT ACTTATACAT AAAGGCTTTA TTAATAATTA
1498  ATTAAATATA CCAGTTTAAT GAAATTTATA CTAATCTTTT TCTAATTTAA TATTATTATG
1558  TGTTTAACTT TAAGTATTCT CCTTAATTTC TTATTGCACT ACTCTCATCT TATCCTCATT
1618  CACTTTTTCT ATATCAAATT TTAATTTTTT TCCTACTTTC TTTCCATTTA AGATTTTTCT
1678  AAATTCCCTT TTGAAAAACC ATAAAAnTTA TAATTTTGCT TATATTTCTT TCATTTCTTA
1738  ATTGTTAAAA ACTCAAATTT GATTTTAACA TGAAATTTTT CCTCTCTTAA ATATTAAATT
1798  CAATATTAAA ATTGAGTAAA ATAGCGCATT TTTGCTTATT TAATAATTAT GATTGTAGTT
1858  TATAATTGAA AATCGAGTGA TTACAAATTC TTATTGAAAC TAGAATATTT AATTATTTAA
1918  ATAATAATAT TCAAAAGTAT TTTAAACCTA ATTTAAGAAT TAAATCTTTT ATTAGTATTT
1978  ACCATTTAAG AATAAATTAG TCTTAAAAAT ATAAATTTCA TAATGGTCAT TAAAAATCGC
2038  TATTTAAATC TAATTAAATT GTAAATTAAA AATCTTAATA CTAAAATTAA AAATTTTTAG
2098  AAATGGCTGA GAAAAATAAC AAGAAATAAT TCTAAATTAT AATTGGCTTA GTAGGAAAGA
2158  TC
```

macro- and micronuclei, it was only necessary to screen a library from one of these sources. Phage positive for hybridization to yeast H4 sequences were plaque purified from a Tetrahymena micronuclear DNA library prepared by partial Eco·RI digestion and insertion into λ Charon 4A (see 31 for details). When phage DNA was digested with Eco·RI, all inserts were found to be ~8.6 kb, identical in size to the larger Eco·RI fragment containing the H4-I gene, found in macro- or micronuclear DNA. Insert DNA from one positive phage (λGB508) was digested with HindIII and Eco·RI and hybridized with yeast H4 probe. The H4 hybridization was located on a 4.3 kb Eco·RI-HindIII fragment which was then subcloned into the plasmid vector RVII Δ7 (to give p508.8; Fig. 2). The H4-I coding region was mapped by hybridizing yeast H4 probe to a variety of restriction digests of p508.8 (Fig. 2). The H4 coding region was localized to a single HhaI fragment of about 1 kb.

Sequence of the H4-I Locus

Based on the restriction mapping data, a Sau3A-Sau3A fragment containing the H4-I gene was sub-cloned into M13. A series of DNAseI-shortened subclones was prepared by the method of Hong (37) and sequenced by the method of Sanger et al. (36). The sequencing strategy is summarized in Fig. 3. We have sequenced 2159 nucleotides centering on the 306 bases which code for the H4 protein (Fig. 4). The predicted amino acid sequence matches exactly the published Tetrahymena H4 protein sequence (25,26). Surprisingly, the Tetrahymena H4 coding sequence is 46% G+C, in contrast to the low GC content (< 25%) of the total DNA (41) and of the sequences surrounding H4-I.

The Tetrahymena H4 sequence shares 78.4% homology (80/102 residues) with yeast H4 at the amino acid level and 78.4% homology (240/306 nucleotides) at the nucleotide level (Fig. 5). The only manipulation required to obtain this homology was to delete a yeast arginine codon at amino acid position 3 and a Tetrahymena serine codon between positions 18 and 19 to account for the unique deletion and insertion of these residues found in Tetrahymena H4 (25). When Tetrahymena H4 sequences are similarly compared to those of multicellular animals (sea urchin, Xenopus, mouse), the amino acid sequence homology is similar to that of yeast, ~80%. However, the nucleotide sequence homology between Tetrahymena H4 genes and animal H4s is only 60-65%.

Figure 4. DNA sequence of Tetrahymena H4 and flanking sequences.
To ensure accurate sequencing data, most regions were sequenced in two independent clones and every gel was analyzed independently by two different people. Lower case letters and "n" represent uncertainties in the DNA sequence.

HISTONE H4 AMINO ACID SEQUENCES

TH4–I	ALA	GLY		GLY	LYS	GLY	GLY	LYS	GLY	MET	GLY	LYS	VAL	GLY	ALA
YH4–II	SER	***	ARG	***	***	***	***	***	***	LEU	***	***	GLY	***	***
										10					
	LYS	ARG	HIS	SER	ARG	LYS	SER	ASN	LYS	ALA	SER	ILE	GLU	GLY	ILE
	***	***	***		***	***	ILE	LEU	ARG	ASP	ASN	***	GLN	***	***
						20									
	THR	LYS	PRO	ALA	ILE	ARG	ARG	LEU	ALA	ARG	ARG	GLY	GLY	VAL	LYS
	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***
	30										40				
	ARG	ILE	SER	SER	PHE	ILE	TYR	ASP	ASP	SER	ARG	GLN	VAL	LEU	LYS
	***	***	***	GLY	LEU	***	***	GLU	GLU	VAL	***	ALA	***	***	***
						50									
	SER	PHE	LEU	GLU	ASN	VAL	VAL	ARG	ASP	ALA	VAL	THR	TYR	THR	GLU
	***	***	***	***	SER	***	ILE	***	***	SER	***	***	***	***	***
	60										70				
	HIS	ALA	ARG	ARG	LYS	THR	VAL	THR	ALA	MET	ASP	VAL	VAL	TYR	ALA
	***	***	LYS	***	***	***	***	***	SER	LEU	***	***	***	***	***
						80									
	LEU	LYS	ARG	GLN	GLY	ARG	THR	LEU	TYR	GLY	PHE	GLY	GLY		
	***	***	***	***	***	***	***	***	***	***	***	***	***		
	90										100				

78% HOMOLOGY

HISTONE H4 NUCLEOTIDE SEQUENCES

TH4–I	GCC	GGT		GGT	AAA	GGT	GGT	AAA	GGT	ATG	GGT	AAA	GTC	GGA	GCC
YH4–II	T**	***	AGA	***	***	***	***	***	***	C*A	***	***	*GT	**T	***
										10					
	AAG	AGA	CAC	TCC	AGA	AAG	TCT	AAC	AAG	GCT	TCC	ATT	GAA	GGT	ATT
	***	C*T	***		***	***	AT*	CTA	*GA	*A*	AA*	**C	C**	***	***
						20									
	ACG	AAG	CCC	GCT	ATC	AGA	AGA	TTA	GCT	AGA	AGA	GGT	GGT	GTT	AAG
	**T	***	**A	***	***	***	***	***	***	***	***	***	***	**C	***
	30										40				
	AGA	ATT	TCC	TCT	TTC	ATT	TAC	GAC	GAC	TCC	AGA	CAA	GTC	TTG	AAG
	C*T	***	**T	GG*	**G	**C	***	**A	**A	GT*	***	GCC	***	***	**A
						50									
	TCT	TTC	TTA	GAA	AAC	GTT	GTT	AGA	GAC	GCT	GTC	ACT	TAC	ACT	GAA
	**C	***	**G	***	TC*	**C	A*C	**G	***	T**	**T	***	***	***	***
	60										70				
	CAC	GCT	AGA	AGA	AAA	ACC	GTC	ACT	GCT	ATG	GAC	GTT	GTC	TAC	GCC
	***	**C	*AG	***	**G	**T	**T	***	T**	T**	**T	***	**T	**T	**T
						80									
	CTC	AAG	AGA	CAA	GGC	AGA	ACT	CTC	TAT	GGT	TTC	GGT	GGT		
	T*G	***	***	***	**T	***	**C	T*A	***	***	***	***	***		
	90										100				

78% HOMOLOGY

Figure 5. Comparison of Tetrahymena and yeast H4 coding regions.
Sequences are aligned for maximum homology. * represents a common nucleotide/amino acid.

Sequences Flanking H4 Genes in Macronuclear DNA

The sequences flanking λGB508 were mapped by hybridizing the 0.8 kb EcoR·1-BglII fragment of p508.8 or the HindIII-Eco·RI fragment of λGB508 (subcloned into RVIIΔ7 and designated p508.7) to HhaI, HindIII, Sau3A, HaeIII or TaqI digested macronuclear DNA. The data from these experiments allowed us to map about 20 kb from the leftward EcoR·1 site and about 12 kb from the rightward Eco·RI site (Fig. 2). To determine if the H4-II gene was located on these flanking sequences, macronuclear DNA was digested with HhaI (the farthest restriction sites we have mapped), blotted and probed with yeast H4, with the Eco·RI-Bgl II fragment of p508.8 or with p508.7. As seen in Fig. 6, the other H4 sequence does not co-migrate with the bands produced by hybridization to these flanking sequences. Thus, the H4-II gene is not located within 20 kb to the left or 12 kb to the right of H4-I. When a blot similar to that in Fig. 6 was probed with yeast H3 gene sequence, one hybridization band co-migrated with H4-II. Similarly the H4-II gene co-migrated with an H3 gene on Southern blots of macronuclear DNA digested with either Eco·RI or HindIII (data not shown). These results indicate that H4-II, but not H4-I is closely linked to an H3 gene.

Chromosomal location of Tetrahymena H4 Genes: Nullisomic Mapping

Since the two Tetrahymena H4 genes are not closely linked, they could be on different chromosomes. In Tetrahymena it is possible to determine the chromosomal location of specific genes using nullisomic strains (42) that are missing both copies of one or more micronuclear chromosomes. Such strains are viable because they have a normal macronuclear DNA complement. Chromosomal locations were determined by hybridizing macro- or micronuclear DNAs from strains CU358 (missing chromosomes 3,4 and 5), CU359 (missing 2,3 and 5) and CU383 (missing 4) to probes from either p508.7 (a unique sequence flanking H4-I) or from yeast H4 gene (which hybridizes to both H4-I and H4-II). These strains allowed us to distinguish between all but chromosomes 3 and 5. All DNA samples except CU383 and CU358 micronuclear DNA hybridized strongly to p508.7 (Fig. 7) indicating that H4-I (which is linked to the sequence in p508.7) is on chromosome 4. The yeast H4 probe hybridized less to micronuclear DNA from CU359 and CU383 than to macronuclear DNA and did not hybridize significantly to micronuclear DNA from strain CU358. Since chromosome 4 which carries the H4-1 gene is missing in CU383, significant hybridization to micronuclear DNA from this strain indicates that the H4-II gene must be on a different chromosome. Since CU359 has chromosome 4, reduced hybridization to micronuclear DNA from this strain indicates the H4-II gene

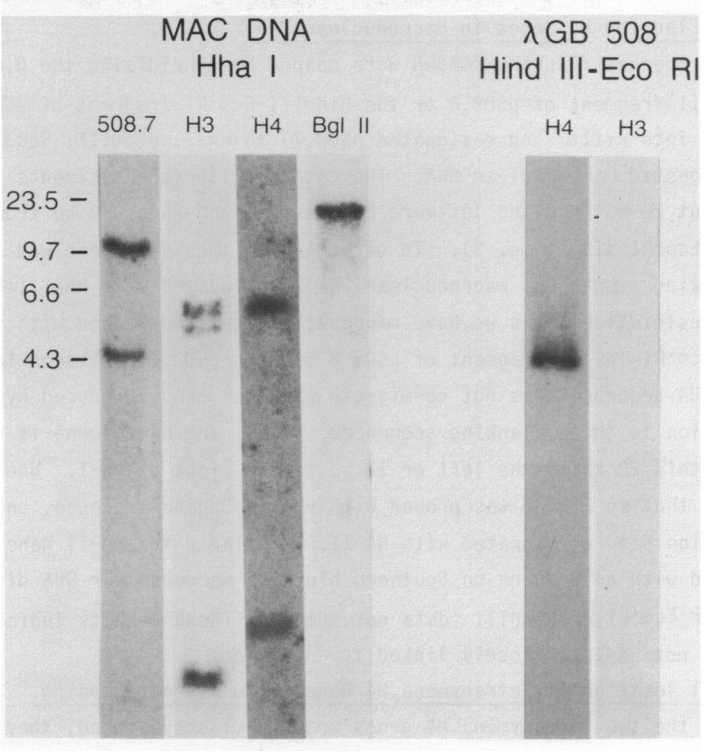

Figure 6. The two Tetrahymena H4 genes are not closely linked.
Hha I digested macronuclear DNA was probed with the yeast H4 or H3 probe or with end fragments from the λGB508 insert (p508.7 or the gel purified RI-BglII subfragment). The p508.7 subclone hybridizes to two bands because a Hha I site exists within this insert. Note that the yeast H4 probe does not hybridize to the same sequences as the end fragments from λGB508. This indicates that H4-II (band migrating at ~6 kb in the H4 lane) is not located on the flanking sequences of H4-I (band migrating at 1 Kb in the H4 lane).

must be on chromosome 2,3 or 5. Absence of hybridization to micronuclear DNA from CU358 narrows the location of H4-II to chromosome 3 or 5. Thus the two H4 genes in Tetrahymena, both of which are presumably active (28) are completely unlinked.

DISCUSSION

Gene Expression in Macronuclei is not Invariably Linked to Gene Rearrangement

Most, if not all genes are transcriptionally inert in the micronucleus and transcriptionally active in the macronucleus of vegetative Tetrahymena (see 41 for review). Recent studies have also shown that rearrangement of the germ line (micronuclear) configuration of many DNA sequences accompanies

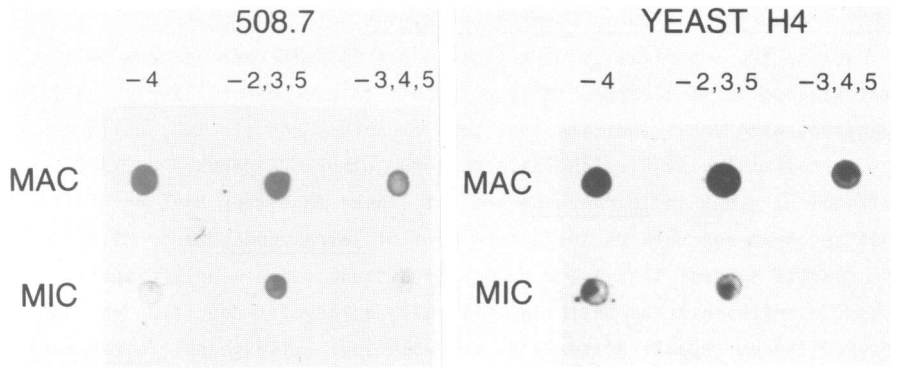

Figure 7. Chromosomal location of Tetrahymena H4 genes.
Macro- and micronuclear DNA from nullisomic strains (see results for details) was probed with p508.7 (a unique sequence which flanks the H4-I gene and therefore serves to determine its location) or with yeast H4 (which hybridizes to both H4 genes). The H4-I gene was localized to chromosome 4 while the H4-II gene was localized to either chromosome 3 or 5. Macronuclear DNA serves as a positive hybridization control. The small amount of hybridization seen in some of the micronuclear spots is due to low levels of macronuclear contamination.

macronuclear differentiation in Tetrahymena, and that at least one gene which must be expressed in macronuclei is adjacent to rearranged sequences (39,40). Since DNA sequence rearrangements in other systems appear to be associated with gene activation (43-45), it is tempting to suggest that the rearrangements in Tetrahymena macronuclei serve the same purpose.

The question arises as to whether all genes which become active in macronuclei must be closely associated with sequence rearrangements. Our sequence analysis of the cloned H4-I gene shows that it codes for a protein with the known sequence of Tetrahymena H4. Previously published studies indicate that there are two independently regulated H4 mRNAs in Tetrahymena (28) suggesting that two H4 genes are active. Therefore, H4-I is, almost certainly, a functional gene. Our comparison of the organization of H4 sequences in macro- and micronuclei indicates that no rearrangements are closely associated with either H4 gene in macronuclei (it should be noted that the Southern blotting techniques used in these comparisons are similar to those which easily demonstrated rearrangements associated with other sequences). While these results do not preclude the possibility that undetectably small rearrangements or distant rearrangements are required for expression of H4-I and H4-II in macronuclei, it is clear that major adjacent rearrangements of a type previously described are not required.

Histone Gene Organization in Diverse Eukaryotes

Because the organization of histone genes in different members of the animal kingdom is so diverse, it is difficult to make generalized comparisons. Comparisons with yeast indicate that both organisms contain two, unlinked H4 genes. Preliminary studies indicate that two other H2A genes and 2-3 additional H3 genes in Tetrahymena are not linked to either H4-I or H4-II. We do not yet have any data on the arrangement of Tetrahymena H2B or H1 genes. These results suggest that a low number of histone genes widely dispersed in the genome represents the basic (primitive?) eukaryotic condition and the clustered tandem repeats arose later in evolution. Like yeast, Tetrahymena histone mRNAs are polyadenylated (28,46) indicating that the absence of poly A tails may also be of recent evolutionary origin. In yeast, both H4 genes are closely linked to H3 genes; in Tetrahymena, we have shown by Southern analysis that the H4-II gene is linked to an H3 gene while the H4-I gene clearly is not. The significance of this difference is not known.

Conservation of Nucleotide Sequences of H4 Genes

The nucleotide sequence coding for Tetrahymena H4 is unusual in a number of respects. The G+C content, 46%, is surprisingly high, especially since it is theoretically possible to construct an H4 gene with the Tetrahymena amino acid composition that is only 32% G+C. The high G+C content of the coding region also contrasts with that of the total genome and the flanking sequences, both of which are less than 25%. Similar limits on the use of AT-rich codons appear to occur in other organisms. Both Oxytricha and Dictyostelium have genomic G+C contents of ~20-25% while the genes sequenced to date (47,48) invariably have protein coding regions with G+C contents closer to 50%. The G+C content of the H4-I region of Tetrahymena changes abruptly outside of the coding region. Since the H4-I transcript is at least 850b (28), while the coding region is only 306b, the high G+C content of the coding sequences suggests that selection pressure against AT-rich base pairs operates at the level of the translational machinery rather than at the level of genome organization or transcription.

Turner and Woodland (49) noted evolutionary constraints on codon usage when comparing H4 sequences from diverse animals. There was frequent use of the same codon in the same position of the H4 molecule for all H4 sequences analyzed. When we extended this analysis to yeast (22) and Tetrahymena H4's, only 2 invariant codons remained: AAG (Lys) at amino acid residue 31 and CAC (His) at position 75. Since Tetrahymena H4 diverges considerably from those of other organisms, conservation of these residues may point to their

essential role in forming stable structures in H4 mRNAs.

Turner and Woodland (49) noted that codons for the conserved serines at positions 1 and 47 were invariably TCX rather than AGPy. This rule applies as well to Tetrahymena and yeast at serine 47 and to yeast at serine 1. However, in Tetrahymena serine 1 is replaced by alanine owing to a single base substitution. It is striking to note that Tetrahymena and yeast H4s contain 7 and 6 serines respectively instead of the 2 commonly found in animal H4s. All serines are encoded by TCX. The codons frequently represent multiple base substitutions from the codons found at the homologous positions in animal H4 genes. Thus, the AGPy codon seems to be universally excluded from histone H4 genes.

Turner and Woodland (49) also observed that in animal H4 genes, the codon ATC is used 52 out of 53 times for isoleucine and TTC is used almost exclusively for phenylalanine. In Tetrahymena ATT is used 4 times while ATC is used only once. In yeast ATC is used 4 of 7 times and ATT the remaining 3. TTC is used invariably for phenylalanine in both yeast and Tetrahymena. Thus, lower eukaryotic H4 genes resemble those of higher organisms in the absence of ATA and TTT codons, but differ from them in that they use the remaining two isoleucine codons indiscriminately.

Tetrahymena and yeast H4 amino acid sequences are ~78% homologous. Given the possibility of silent, third base wobble and even of occasional first and second base changes in codons, one might have expected even more divergence in nucleotide sequence. However, there is 78% homology between the H4 coding regions of yeast and Tetrahymena. When the Tetrahymena H4 sequence is compared to those of higher organisms, the amino acid sequences are ~80% homologous but, as expected, homology at the nucleotide level is considerably lower, 60-65%. These results suggest that selection pressures operate at the level of nucleotide sequences independent of their effects on protein sequence and that the lower eukaryotes may be subject to different selection pressures than higher eukaryotes. Since the sequence homology between yeast and Tetrahymena H4's decreases markedly outside the coding region, these results again suggest that it is some feature of the translational machinery that is responsible for these evolutionary constraints.

Finally, we have noted remarkable homology between the first 27 nucleotides of Tetrahymena H4 and yeast H2A coding regions (50; beginning with the initiator ATG). Twenty five of the first 27 nucleotides are exact matches. Differences in a single nucleotide (amino acid position 2, T→G; amino acid position 9, G→C) result in different amino acids being encoded at

those positions when Tetrahymena H4 and yeast H2A are compared. When sequences coding for the first 9 amino acids of yeast H2A and yeast H4 are compared, 26 of the first 27 nucleotides are exact matches. The only manipulation required to obtain this homology was to delete the yeast H4 arginine codon at amino acid position 3. This manipulation was also required for comparison of Tetrahymena and yeast H4s. Similar slight manipulations of single residues serve to align the first 9 residues of the amino termini of H4s and H2As of higher organisms as well. The functional and evolutionary sigificance of these observations is unclear. However, it is interesting to speculate that histone H4 and H2A may have originally arisen from a single polypeptide which later diverged. Clearly, some functional constraint has resulted in retention of homology at the amino terminus.

ACKNOWLEDGEMENTS

We thank M. M. Smith for supplying the yeast histone clone and a detailed restriction map and P. J. Bruns for supplying Tetrahymena nullisomic strains. We acknowledge the technical assistance of Paul Tucci and James Cook. This work was supported by research grants from the National Institutes of Health to M.A.G. and M.C.Y. G.A.B. is a postdoctoral fellow of the American Cancer Society.

REFERENCES

1. Maxson, R., R. Cohn and L. Kedes. (1983) Ann. Rev. Genet. 17: 239-277.
2. Lifton, R.P., M.L. Goldberg, R.W. Karp and D.S. Hogness. (1977) Cold Spring Harbor Symp. Quant. Biol. 42: 1047-1051.
3. Goldberg, M.L. (1979) Ph.D. thesis, Stanford University, Palo Alto, California.
4. Gross, K., W. Schaffner, J. Telford and M. Birnstiel. (1976) Cell 8: 479-484.
5. Birnstiel, M.L., R. Portmann, M. Busslinger, W. Schaffner, E. Probst and A. Kressmann. (1979) Alfred Benzon Symp. 13: 117-129.
6. Busslinger, M., R. Portmann, and M.L. Birnstiel. (1979) Nuc. Acids Res. 6: 2997.
7. Kedes, L.H. (1979) In Annual Review of Biochemistry 48: 837-870.
8. Overton, G.C. and E.S. Weinberg. (1978) Cell 14: 247-257.
9. Cohn, R.H. and L.H. Kedes. (1979a) Cell 18: 843-853.
10. Cohn, R.H. and L.H. Kedes. (1979b) Cell 18: 855-864.
11. Old, R.W., H.R. Woodland, J.E.M. Ballantine, T.C. Aldridge, C.A. Newton, W.A. Bains and P.C. Turner. (1982) Nuc. Acids. Res. 10: 7561-7580.
12. Moorman, A.F.M., R.T.M. de Laaf and O.H.J. Destree. (1980) Gene 10(3), 185-193.
13. van Dongen, W., L. de Laaf, R. Zaal, A. Moorman, A., D. Destree. (1981) Nucleic Acids Research 9(10): 2297-2311.
14. Ruberti, I., P. Fragapane, P. Pierandrei-Amaldi, E. Beccari, F. Almadi and I. Bozzoni. (1982) Nuc. Acids Res. 10: 7543-7559.
15. Zernik, M., N. Heintz, I. Boime and R. Roeder. (1980) Cell 22: 807-815.
16. Sugarman, B. J., J. Dodgson and J. D. Engel. (1983) J. Biol. Chem.

258: 9005-9016.
17. Engel, J. and J. Dodgson. (1981) Proc. Natl. Acad. Sci. USA 78: 2856-2860.
18. Seiler-Tuyns, A. and M. Birnstiel. (1981) J. Mol. Bio. 151: 607-625.
19. Stephenson, E.C., H.P. Erba and J.G. Gall. (1981) Cell 24: 639-647.
20. Childs, G., C. Nocente-McGrath, T. Lieber, C. Holt and J.A. Knowles. (1982) Cell 31: 383-393.
21. Hereford, L., K. Fahrner, J. Woolford, Jr., M. Rosbash and D. B. Kaback. (1979) Cell 18: 1261-1271.
22. Smith, M.M. and O.S. Andresson. (1983) J. Mol. Biol. (in press).
23. Tabata, T., K. Sasaki and M. Iwabuchi. (1983) Nuc. Acids Res. 11: 5865-5875.
24. Elsevier, S.M., H.S. Lipps and G. Steinbruck. (1979) Chromosoma 69: 291-306.
25. Glover, C.V.C. and M. A. Gorovsky. (1979) Proc. Natl. Acad. Sci. USA 76: 585-589.
26. Hayashi, H., M. Nomoto and K. Iwai. (1980) Proc. Japan Acad. 56 Ser B. 579-584.
27. Gorovsky, M.A., M.C. Yao, J.B. Keevert, and G.L. Pleger. (1975) Methods in Cell Biology. P.M. Prescot (ed.) Academic Press, Inc., New York, IX: 311-327.
28. Bannon, G.A., F.J. Calzone, J.K. Bowen, C.D. Allis, and M.A. Gorovsky. (1983) Nucleic Acids Res. 11: 3903-3917.
29. Maniatis, T., A. Jeffrey and D.G. Kleid. (1975) Proc. Natl. Acad. Sci. USA 72: 1184-1188.
30. Southern, E.M. (1975) J. Mol. Biol. 98: 503-517.
31. Yao, M-C. (1981) Cell 24: 765-774.
32. Benton, W.D. and R. Davis. (1977) Science, 196: 180-182.
33. Blattner, F. R., A. E. Blechl, K. Denniston-Thompson, H. E. Faber, J. E. Richards, J. L. Slightom, P. W. Tucker and O. Smithies. (1978) Science 202: 1279-1283.
34. Lynn, D. A., L. M. Angerer, A. M. Bruskin, W. H. Klein and R. C. Angerer. (1983) Proc. Natl. Acad. Sciences USA 80:2656-2660.
35. Maniatis, T., E.F. Fritsch and J. Sambrook. (1982) Cold Spring Harbor Laboratory Publication.
36. Sanger, F., S. Nicklen, and A. R. Coulson. (1977) Proc. Natl. Acad. Sci. USA 74: 5463-5467.
37. Hong, G.F. (1982) J. Mol. Biol. 158: 539-549.
38. Messing, J., R. Crea and P.H. Seeburg. (1981) Nucleic Acids Res. 9: 309-321.
39. Yao, M.-C., J. Choi, S. Yokoyama, C. Austerberry and C.-H. Yao (1984) Cell, In Press.
40. Callahan, R. C., G. Shalke and M. A. Gorovsky. (1984) Cell, In Press.
41. Gorovsky, M. A. (1980) Annual Rev. of Genetics 14: 203-239.
42. Bruns, P.J. and T.E.B. Brussard. (1981) Science 213: 549-551.
43. Tonegawa, S. (1983) Nature 302: 575-581.
44. Borst, P. and G. A. M. Cross. (1982) Cell 29: 291-303.
45. Nasmyth, K. A. (1982) Annual Rev. of Genetics 16: 439-500.
46. Fahrner, K., J. Yarger and L. Hereford. (1980) Nuc. Acids Res. 8: 5725-5737.
47. Kimmel, A.R. and R.A. Firtel. (1983) Nuc. Acids Res. 11: 541-552.
48. Kaine, B. P. and B. B. Spear. (1982) Nature 295: 430-432.
49. Turner, P.C. and H.R. Woodland. (1982) Nuc. Acids Res. 10: 3769-3780.
50. Choe, J., D. Kolodrubetz, and M. Grunstein. (1982) Proc. Natl. Acad. Sci. USA 79: 1484-1487.